



RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/608,713

DATE: 06/28/2002

TIME: 14:13:56

Input Set : D:\seq list.txt

Output Set: N:\CRF3\06282002\I608713.raw

RECEIVED

JUL 08 2002

TECH CENTER 1600/2900

4 <110> APPLICANT: Ago, Hideo
 5 Miyano, Masashi
 6 Adachi, Tsuyoshi
 8 <120> TITLE OF INVENTION: HCV Polymerase Suitable for Crystal
 9 Structure Analysis and Method for Using the Enzyme
 12 <130> FILE REFERENCE: SHIM007
 14 <140> CURRENT APPLICATION NUMBER: 09/608,713
 C--> 15 <141> CURRENT FILING DATE: 2002-06-30
 17 <150> PRIOR APPLICATION NUMBER: 11-188630
 18 <151> PRIOR FILING DATE: 1999-07-02
 20 <150> PRIOR APPLICATION NUMBER: 11-192488
 21 <151> PRIOR FILING DATE: 1999-07-07
 23 <160> NUMBER OF SEQ ID NOS: 13
 25 <170> SOFTWARE: FastSEQ for Windows Version 4.0
 27 <210> SEQ ID NO: 1
 28 <211> LENGTH: 591
 29 <212> TYPE: PRT
 30 <213> ORGANISM: Hepatitis C Virus
 32 <400> SEQUENCE: 1
 33 Ser Met Ser Tyr Thr Trp Thr Gly Ala Leu Ile Thr Pro Cys Ala Ala
 34 1 5 10 15
 35 Glu Glu Ser Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg
 36 20 25 30
 37 His His Asn Met Val Tyr Ala Thr Ser Arg Ser Ala Gly Leu Arg
 38 35 40 45
 39 Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Asp His Tyr
 40 50 55 60
 41 Arg Asp Val Leu Lys Glu Met Lys Ala Lys Ala Ser Thr Val Lys Ala
 42 65 70 75 80
 43 Lys Leu Leu Ser Val Glu Glu Ala Cys Lys Leu Thr Pro Pro His Ser
 44 85 90 95
 45 Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Asn Leu Ser
 46 100 105 110
 47 Ser Lys Ala Val Asn His Ile His Ser Val Trp Lys Asp Leu Leu Glu
 48 115 120 125
 49 Asp Thr Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val
 50 130 135 140
 51 Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile
 52 145 150 155 160
 53 Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr
 54 165 170 175
 55 Asp Val Val Ser Thr Leu Pro Gln Val Val Met Gly Ser Ser Tyr Gly
 56 180 185 190

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/608,713

DATE: 06/28/2002

TIME: 14:13:57

Input Set : D:\seq list.txt

Output Set: N:\CRF3\06282002\I608713.raw

```

57 Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Asn Thr Trp
58      195                      200                      205
59 Lys Ser Lys Lys Asn Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe
60      210                      215                      220
61 Asp Ser Thr Val Thr Glu Asn Asp Ile Arg Val Glu Glu Ser Ile Tyr
62 225                      230                      235                      240
63 Gln Cys Cys Asp Leu Ala Pro Glu Ala Arg Gln Ala Ile Lys Ser Leu
64      245                      250                      255
65 Thr Glu Arg Leu Tyr Ile Gly Gly Pro Leu Thr Asn Ser Lys Gly Glu
66      260                      265                      270
67 Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser
68      275                      280                      285
69 Cys Gly Asn Thr Leu Thr Cys Tyr Leu Lys Ala Ser Ala Ala Cys Arg
70      290                      295                      300
71 Ala Ala Lys Leu Gln Asp Cys Thr Met Leu Val Asn Gly Asp Asp Leu
72 305                      310                      315                      320
73 Val Val Ile Cys Glu Ser Ala Gly Thr Gln Glu Asp Ala Ala Ser Leu
74      325                      330                      335
75 Arg Val Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp
76      340                      345                      350
77 Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser
78      355                      360                      365
79 Asn Val Ser Val Ala His Asp Ala Ser Gly Lys Arg Val Tyr Tyr Leu
80      370                      375                      380
81 Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala
82 385                      390                      395                      400
83 Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Tyr Ala
84      405                      410                      415
85 Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Ile
86      420                      425                      430
87 Leu Leu Ala Gln Glu Gln Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr
88      435                      440                      445
89 Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Gln Ile Ile Glu
90      450                      455                      460
91 Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly
92 465                      470                      475                      480
93 Glu Ile Asn Arg Val Ala Ser Cys Leu Arg Lys Leu Gly Val Pro Pro
94      485                      490                      495
95 Leu Arg Val Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu
96      500                      505                      510
97 Ser Gln Gly Gly Arg Ala Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp
98      515                      520                      525
99 Ala Val Lys Thr Lys Leu Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln
100     530                      535                      540
101 Leu Asp Leu Ser Gly Trp Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile
102 545                      550                      555                      560
103 Tyr His Ser Leu Ser Arg Ala Arg Pro Arg Trp Phe Met Leu Cys Leu
104      565                      570                      575
105 Leu Leu Leu Ser Val Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/608,713

DATE: 06/28/2002

TIME: 14:13:57

Input Set : D:\seq list.txt

Output Set : N:\CRF3\06282002\I608713.raw

```

106          580          585          590
109 <210> SEQ ID NO: 2
110 <211> LENGTH: 2889
111 <212> TYPE: DNA
112 <213> ORGANISM: Artificial Sequence
114 <220> FEATURE:
115 <221> NAME/KEY: CDS
116 <222> LOCATION: (1)...(1743)
117 <223> OTHER INFORMATION: DNA encoding fusion protein consistin of a portion
118       of HCV polymerase and histidine tag at the
119       C-terminus
121 <221> NAME/KEY: misc_feature
122 <222> LOCATION: 144, 152, 214, 294, 312, 560, 622, 716, 776, 939, 1007, 1017,
123       1031, 1095, 1183, 1249, 1270, 1349, 1405, 1408, 1501, 1588,
124       1604, 1679, 1824, 1899, 2076, 2086, 2221, 2225, 2240, 2313,
125       2445, 2605, 2634, 2760
126 <223> OTHER INFORMATION: n = A,T,C or G
W--> 128 <221> misc_feature
129 <222> LOCATION: 144, 152, 1679, 1824, 1899, 2076, 2086, 2221, 2225, 2240,
130       2313, 2445, 2605, 2634, 2760
131 <223> OTHER INFORMATION: n = A,T,C or G
133 <223> OTHER INFORMATION: DNA encoding fusion protein consisting of a
134       portion of HCV polymerase and histidine tag at the
135       C-terminus
W--> 137 <221> NAME/KEY: misc_feature
W--> 138 <222> LOCATION: 144, 152, 214, 294, 312, 560, 622, 716, 776, 939, 1007, 1017,
W--> 139       1031, 1095, 1183, 1249, 1270, 1349, 1405, 1408, 1501, 1588,
W--> 140       1604, 1679, 1824, 1899, 2076, 2086, 2221, 2225, 2240, 2313,
W--> 141       2445, 2605, 2634, 2760
W--> 142 <223> n = A,T,C or G
W--> 144 <221> NAME/KEY: misc_feature
W--> 145 <222> LOCATION: 144, 152, 214, 294, 312, 560, 622, 716, 776, 939, 1007, 1017,
W--> 146       1031, 1095, 1183, 1249, 1270, 1349, 1405, 1408, 1501, 1588,
W--> 147       1604, 1679, 1824, 1899, 2076, 2086, 2221, 2225, 2240, 2313,
W--> 148       2445, 2605, 2634, 2760
W--> 149 <223> n = A,T,C or G
W--> 151 <400> 2
152 atg tca atg tcc tac aca tgg aca ggc gcc ttg atc acg cca tgc gct 48
153 Met Ser Met Ser Tyr Thr Trp Thr Gly Ala Leu Ile Thr Pro Cys Ala
154 1 5 10 15
155 mts rmt srt yrt hrt rth rgy aaU thr rcy saa gcg gag gaa agc aaq 96
W--> 157 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Asn Xaa Xaa Xaa Ala Glu Glu Ser Lys
158 20 25 30
W--> 160 ctg ccc atc aac gcg ttg agc aac tct ttg ctg aag ugu sry sur asn 144
W--> 161 Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Lys Cys Xaa Xaa Xaa
162 35 40 45
W--> 164 aaU sra sns ruu cgc cac cat aac atg gtt tat gcc aca aca tct cgc 192
W--> 165 Asn Xaa Xaa Xaa Arg His His Asn Met Val Tyr Ala Thr Thr Ser Arg
166 50 55 60

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/608,713

DATE: 06/28/2002

TIME: 14:13:57

Input Set : D:\seq list.txt

Output Set: N:\CRF3\06282002\I608713.raw

```

W--> 168 agc gca ggc ctg arg hsh sas nmt vat yra ath rth rsr arg sra agy      240
W--> 169 Ser Ala Gly Leu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
170 65 70 75 80
172 ucg gca gaa gaa ggt cac ctt tga cag act gca agt cct gga cga cca      288
173 Ser Ala Glu Glu Gly His Leu * Gln Thr Ala Ser Pro Gly Arg Pro
174 85 90 95
W--> 176 car ggn ysy sva thr has arg ugn vau asa shs tac cgg gac gtg ctc      336
W--> 177 Gln Gly Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Tyr Arg Asp Val Leu
178 100 105 110
180 aag gag atg aag gcg aag gcg tcc aca gtt aag tyr arg asv auy sgu      384
W--> 181 Lys Glu Met Lys Ala Lys Ala Ser Thr Val Lys Xaa Xaa Xaa Xaa Xaa
182 115 120 125
184 mty saa ysa asr thr vay sgc taa act cct atc cgt aga gga agc ctg      432
W--> 185 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa * Thr Pro Ile Arg Arg Gly Ser Leu
186 130 135 140
188 caa gct gac gcc ccc aca taa ysu usr vag ugu aac ysy sut hrr rhs      480
W--> 189 Gln Ala Asp Ala Pro Thr * Xaa Xaa Xaa Cys Asn Xaa Xaa Xaa Xaa
190 145 150 155
192 tgg gcc aaa tcc aag ttt ggc tat ggg gca aag gac gtc cgg aac cta      528
193 Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Asn Leu
194 160 165 170
W--> 196 sra ays sry shg yty rgy aay sas vaa rga snu tcc agc aag gcc gtt      576
W--> 197 Xaa Xaa Xaa Xaa Xaa Xaa Asn Xaa Xaa Xaa Xaa Ser Ser Lys Ala Val
198 175 180 185
W--> 200 aac cac atc cac tcc gtg tgg aag gac ttg ctg srs rys aav aas nhs      624
W--> 201 Asn His Ile His Ser Val Trp Lys Asp Leu Leu Xaa Xaa Xaa Xaa Xaa
202 190 195 200 205
204 hss rva try sas uug aag aca ctg tga cac caa ttg aca cca cca tca      672
W--> 205 Xaa Xaa Xaa Xaa Leu Lys Thr Leu * His Gln Leu Thr Pro Pro Ser
206 210 215 220
W--> 208 tgg caa aaa atg agg uas thr vat hrr ast hrt hrm taa ysa sng ugt      720
W--> 209 Trp Gln Lys Met Arg Xaa Xaa Xaa Xaa Xaa Xaa Xaa * Xaa Xaa Xaa
210 225 230 235
212 ttt ctg tgt cca acc aga gaa agg agg ccg taa gcc agc ccg cct tva      768
W--> 213 Phe Leu Cys Pro Thr Arg Glu Arg Arg Pro * Ala Ser Pro Pro Xaa
214 240 245 250
W--> 216 hcy sva gnr guy sgy gya rgy sra aar gua tgg tat tcc cag atc tgg      816
W--> 217 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Lys Val Ser Tyr Ser Gln Ile Trp
218 255 260 265
220 gag tcc gtg tat gcg aga aga tgg ccc tcv ahr asu gyv aar gva cys      864
W--> 221 Glu Ser Val Tyr Ala Arg Arg Trp Pro Xaa Xaa Xaa Xaa Lys Xaa Xaa
222 270 275 280
224 guy smt aaU tat gat gtg gtc tcc acc ctt cct cag gtc gtg atg ggc      912
W--> 225 Xaa Xaa Asn Tyr Asp Val Val Ser Thr Leu Pro Gln Val Val Met Gly
226 285 290 295
W--> 228 tcc tca tac tyr asv ava srt hru rgn vav amt gys rsr tyr gga ttc      960
W--> 229 Ser Ser Tyr Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Gly Phe
230 300 305 310
W--> 232 cag tac tct cct ggg cag cga gtc gag ttc ctg gtg aat acc gyh gnt      1008

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/608,713

DATE: 06/28/2002

TIME: 14:13:57

Input Set : D:\seq list.txt

Output Set: N:\CRF3\06282002\I608713.raw

```

W--> 233 Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Asn Thr Xaa Xaa
      234 315                      320                      325                      330
W--> 236 yrs rrg ygn arg vag uhu vaa snt hrt gga aat caa aga aaa acc cca 1056
W--> 237 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Gly Asn Gln Arg Lys Thr Pro
      238                      335                      340                      345
W--> 240 tgg gct ttt cat atg aca ctc gct gtt rys sry sys asn rmt gyh srt 1104
W--> 241 Trp Ala Phe His Met Thr Leu Ala Val Xaa Xaa Xaa Xaa Xaa Xaa Xaa
      242                      350                      355                      360
      244 yra sth rar gcy stt cga ctc aac ggt cac cga gaa cga cat ccg tgt 1152
W--> 245 Xaa Xaa Xaa Xaa Xaa Arg Leu Asn Gly His Arg Glu Arg His Pro Cys
      246                      365                      370                      375
W--> 248 tga gga gtc aat tha ssr thr vat hrg uas nas arg vag ugu srt acc 1200
W--> 249 * Gly Val Asn Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Cys Xaa Thr
      250                      380                      385                      390
      252 aat gtt gtg act tgg ccc ccg aag cca gac agg cca taa aat cgt yrg 1248
W--> 253 Asn Val Val Thr Trp Pro Pro Lys Pro Asp Arg Pro * Asn Arg Xaa
      254                      395                      400                      405
W--> 256 ncy scy sas uaa rgu aaa rgg naa yss rct cac aga gcg gct tta tat 1296
W--> 257 Xaa Xaa Xaa * Xaa Lys Xaa Xaa Xaa Xaa His Arg Ala Ala Leu Tyr
      258                      410                      415                      420
      260 cgg ggg tcc tct gac taa ttc aaa agg gut hrg uar gut yrg ygy rut 1344
W--> 261 Arg Gly Ser Ser Asp * Phe Lys Arg Arg Xaa Xaa Xaa Xaa Xaa Xaa
      262                      425                      430                      435
W--> 264 hra sns rys gyc aga act gcg gtt atc gcc ggt gcc gcg cga gcg gcg 1392
W--> 265 Xaa Xaa Xaa Xaa Arg Thr Ala Val Ile Ala Gly Ala Ala Arg Ala Ala
      266                      440                      445                      450
W--> 268 tgc tga cga ctg nas ncy sgy tyr arg arg cys arg aas rgy vau thr 1440
W--> 269 Cys * Arg Leu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
      270 455                      460                      465
      272 thr agc tgc ggt aac acc ctc aca tgt tac ttg aag gcc tct gca gcc 1488
W--> 273 Xaa Ser Cys Gly Asn Thr Leu Thr Cys Tyr Leu Lys Ala Ser Ala Ala
      274 470                      475                      480                      485
W--> 276 tgt src ysg yas nth rut hrc yst yru ysa asr aaa acy scg agc tgc 1536
W--> 277 Cys Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Lys Xaa Xaa Ser Cys
      278                      490                      495                      500
      280 gaa gct cca gga ctg cac gat gct cgt gaa cgg aga cga car gaa aay 1584
      281 Glu Ala Pro Gly Leu His Asp Ala Arg Glu Arg Arg Arg Gln Glu Asn
      282                      505                      510                      515
W--> 284 sug nas cys thr mtu vaa sng yas asc tcg tcg tta tct gtg aaa gcg 1632
W--> 285 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Ser Ser Leu Ser Val Lys Ala
      286                      520                      525                      530
W--> 288 cgg gaa ccc aag agg acg cgg cga gcu vav acy sgu sra agy thr gng 1680
W--> 289 Arg Glu Pro Lys Arg Thr Arg Arg Ala Xaa Xaa Xaa Xaa Xaa Xaa Xaa
      290                      535                      540                      545
      292 uas aaa asr cta cga gtc ttc acg gag gct atg act agg tac tcc gcc 1728
W--> 293 Xaa Lys Xaa Leu Arg Val Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala
      294 550                      555                      560                      565
      296 ccc ccc ggg uar gva hthrguaamt thrargtyrs raarrgygac ccgccccaac 1783
W--> 297 Pro Pro Gly Xaa Xaa

```

RAW SEQUENCE LISTING ERROR SUMMARY
 PATENT APPLICATION: US/09/608,713

DATE: 06/28/2002
 TIME: 14:13:58

Input Set : D:\seq list.txt
 Output Set: N:\CRF3\06282002\I608713.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:2; N Pos. 144,152,214,294,312,560,622,716,776,939,1007,1017,1031,1095
 Seq#:2; N Pos. 1183,1249,1270,1349,1405,1408,1501,1588,1604,1679,1824,1899
 Seq#:2; N Pos. 2076,2086,2221,2225,2240,2313,2445,2605,2634,2760
 Seq#:2; Xaa Pos. 17,18,19,20,21,22,23,25,26,27,46,47,48,50,51,52,69,70,71
 Seq#:2; Xaa Pos. 72,73,74,76,77,78,79,80,99,100,101,102,103,104,105,106,107
 Seq#:2; Xaa Pos. 124,125,126,127,128,129,130,131,132,133,134,135,152,153
 Seq#:2; Xaa Pos. 154,157,158,159,160,177,178,179,180,181,182,184,185,186
 Seq#:2; Xaa Pos. 187,204,205,206,207,208,209,210,211,212,230,231,232,233
 Seq#:2; Xaa Pos. 234,235,236,238,239,240,256,257,258,259,260,261,262,263
 Seq#:2; Xaa Pos. 264,282,283,284,285,287,288,289,290,308,309,310,311,312
 Seq#:2; Xaa Pos. 313,314,315,316,317,318,335,336,337,338,339,340,341,342
 Seq#:2; Xaa Pos. 343,344,345,362,363,364,365,366,367,368,369,370,371,372
 Seq#:2; Xaa Pos. 373,389,390,391,392,393,394,395,396,397,399,416,417,418
 Seq#:2; Xaa Pos. 419,421,423,424,425,426,443,444,445,446,447,448,449,450
 Seq#:2; Xaa Pos. 451,452,469,470,471,472,473,474,475,476,477,478,479,480
 Seq#:2; Xaa Pos. 481,498,499,500,501,502,503,504,505,506,507,509,510,529
 Seq#:2; Xaa Pos. 530,531,532,533,534,535,536,537,554,555,556,557,558,559
 Seq#:2; Xaa Pos. 560,561,563,580,581
 Seq#:3; Xaa Pos. 17,18,19,20,21,22,23,25,26,27,46,47,48,50,51,52,69,70,71
 Seq#:3; Xaa Pos. 72,73,74,76,77,78,79,80,98,99,100,101,102,103,104,105,106
 Seq#:3; Xaa Pos. 123,124,125,126,127,128,129,130,131,132,133,134,149,150
 Seq#:3; Xaa Pos. 151,154,155,156,157,174,175,176,177,178,179,181,182,183
 Seq#:3; Xaa Pos. 184,201,202,203,204,205,206,207,208,209,226,227,228,229
 Seq#:3; Xaa Pos. 230,231,232,233,234,235,250,251,252,253,254,255,256,257
 Seq#:3; Xaa Pos. 258,276,277,278,279,281,282,283,284,302,303,304,305,306
 Seq#:3; Xaa Pos. 307,308,309,310,311,312,329,330,331,332,333,334,335,336
 Seq#:3; Xaa Pos. 337,338,339,356,357,358,359,360,361,362,363,364,365,366
 Seq#:3; Xaa Pos. 367,382,383,384,385,386,387,388,389,390,392,408,409,410
 Seq#:3; Xaa Pos. 411,412,414,415,416,417,433,434,435,436,437,438,439,440
 Seq#:3; Xaa Pos. 441,442,458,459,460,461,462,463,464,465,466,467,468,469
 Seq#:3; Xaa Pos. 470,487,488,489,490,491,492,493,494,495,496,498,499,518
 Seq#:3; Xaa Pos. 519,520,521,522,523,524,525,526,543,544,545,546,547,548
 Seq#:3; Xaa Pos. 549,550,552,569,570
 Seq#:12; Xaa Pos. 1,8